Predicting species emergence in simulated complex pre-biotic networks

Markovitch, Omer; Krasnogor, Natalio

Published in:
PLoS ONE

DOI:
10.1371/journal.pone.0192871

PLOS ONE

RESEARCH ARTICLE

Predicting species emergence in simulated complex pre-biotic networks

Omer Markovitch[¤a¤b], **Natalio Krasnogor***

Interdisciplinary Computing and Complex Bio-Systems research group, School of Computing Science, Newcastle University, Newcastle upon Tyne, United-Kingdom

¤a Current Address: Center for Systems Chemistry, Stratingh Institute, University of Groningen, Groningen, The Netherlands
¤b Current Address: Blue Marble Space Institute of Science, Seattle, Washington, United States of America
* natalio.krasnogor@newcastle.ac.uk

Citation: Markovitch O, Krasnogor N (2018) Predicting species emergence in simulated complex pre-biotic networks. PLoS ONE 13(2): e0192871. https://doi.org/10.1371/journal.pone.0192871

Editor: Ricard V. Solé, Santa Fe Institute, SPAIN

Received: November 29, 2017

Accepted: January 31, 2018

Published: February 15, 2018

Data Availability Statement: All data and code are available here: http://ico2s.org/data/extras/gard/; additionally, data have been uploaded to Zenodo and are accessible using the following DOI: 10.5281/zenodo.56534.

Funding: This work was funded by the UK's Engineering and Physical Sciences Research Council (EPSRC) under projects (EP/J004111/2) "Towards a Universal Biological-Cell Operating System (AUdACiOuS)" and (EP/N031962/1) "Synthetic Portabolomics: Leading the way at the crossroads of the Digital and the Bio Economies" (NK). The funders had no role in study design, data

Abstract

An intriguing question in evolution is what would happen if one could "replay" life's tape. Here, we explore the following hypothesis: when replaying the tape, the details ("decorations") of the outcomes would vary but certain "invariants" might emerge across different life-tapes sharing similar initial conditions. We use large-scale simulations of an *in silico* model of pre-biotic evolution called GARD (Graded Autocatalysis Replication Domain) to test this hypothesis. GARD models the temporal evolution of molecular assemblies, governed by a rates matrix (i.e. network) that biases different molecules' likelihood of joining or leaving a dynamically growing and splitting assembly. Previous studies have shown the emergence of so called compotypes, i.e., species capable of replication and selection response. Here, we apply networks' science to ascertain the degree to which invariants emerge across different life-tapes under GARD dynamics and whether one can predict these invariant from the chemistry specification alone (i.e. GARD's rates network representing initial conditions). We analysed the (complex) rates' network communities and asked whether communities are related (and how) to the emerging species under GARD's dynamic, and found that the communities correspond to the species emerging from the simulations. Importantly, we show how to use the set of communities detected to predict species emergence without performing any simulations. The analysis developed here may impact complex systems simulations in general.

Introduction

The Origins of Life (OOL) field attempts to understand the transition from a mixture of lifeless molecules to life-full entities, with protocells [1–4] as intermediate (potentially viable) milestones along the non-living to living spectrum [5]. A widely accepted definition of minimal life is: a self-sustaining system capable of undergoing Darwinian evolution [6], while other definitions are often similar (e.g. [7]). A minimally living entity needs not be a cell as we know it but could be a much simpler protocell [2, 8–15], i.e. container with some necessary molecular content. Two major schools tackle the problem of transition from non-life to life: the

collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

genetic, or replicator-first approach, and the metabolism-first approach. The replicator-first approach focuses on a single self-perpetuating informational biopolymer, e.g., RNA, as the first step, and it is thus often referred to as the "RNA world" [16–20]. In contrast, the metabolism-first approach [2, 9, 11, 21–23] focuses on a network of chemical reactions among simpler chemical components that became endowed with some reproductive characteristics [2, 8, 9, 11–13].

The RNA world, a widely accepted replicator-first scenario, assumes that a molecule similar or analogous to present day RNA played the role of the self-perpetuating biopolymer [17–19, 24]. The mixture of such molecules is assumed to have later evolved both a metabolic network and an encompassing container. The RNA-world draws from RNA's capability to store (sequential) information and certain catalytic activities typical of metabolism [25–28].

The metabolism-first scenario, on the other hand, suggests that the very first life precursors are likely to have been relatively elaborate molecular networks of much simpler organic molecules, thus trading the complexity of the building blocks (e.g. RNA) for the complexity at the ensemble level. One of the first suggested possible chemical pathway for the emergence of life was made by Oparin, who proposed that it could be manifested by the molecular reactions of relatively simple organic molecules in the primordial soup, interacting with each other to spontaneously form colloidal molecular assemblies (coacervates) [8, 29, 30].

The lipid world scenario for OOL is a variant of the metabolism first scenario, which considers a complex chemical system consisting of mixture of mutually interacting simple molecule types which spontaneously form noncovalent assemblies [22, 31]. Importantly, these assemblies store information in the form of non-random molecular compositions–compositional information (i.e. the specific ratio of different molecule types that make up the assembly)–and pass it to progeny via homeostatic growth accompanied by fission. This information transmission is a function similar to what can be done with sequence-based biopolymers such as RNA/DNA/PNA, except that in this case it is compositional information that is preserved and propagated rather than sequential information. Specifically, compositional replication is the transfer (at least partially) of compositional information from parent to progeny, where the process of information transfer is itself a function of the compositional information in the parent entity [32]. The composition encoded in several chemical systems has been shown to affect their physical properties (i.e. phenotypes), supporting the realism of the lipid world. For example, vesicles' lipid-composition has been shown to affect dye encapsulation efficiency [33] or vesicle's structure [34], and genetic programming ("evolutionary algorithms") has been applied to evolve vesicles' formulation [35, 36]. More recently it has been suggested that vesicles can "osmotically" couple otherwise decoupled chemical reactions [37].

The GARD kinetic model is a physio-chemical simulator within the lipid world scenario [31, 38–40]. The model is based on a matrix (named β) that determines the interactions between different molecular types while the system is kept away from thermodynamic equilibrium by assembly fission (Fig 1). GARD dynamics exhibit quasi-stationary states, which appear in the simulation as faithfully replicating molecular assemblies, termed composomes (for compositional genomes) [38]. Clusters of compositionally-similar composomes are called compotypes (for composome types) [41]. These compotypes have been shown to respond to selection [40], exhibit ecology-like population dynamics [42] and exhibit quasispecies behavior including error-catastrophe-like transition [32] and hence have been interpreted as (emergent) species.

The next paragraph presents a more elaborate discussion of selection in GARD, which can be summarised as the following: GARD simulations show compotypes (but note that not every composition is a compotype), these compotypes can respond to external selection (but not always) by changing their frequencies within a population. Under very small alphabet size and very small assembly size this change in frequency seems negligible.

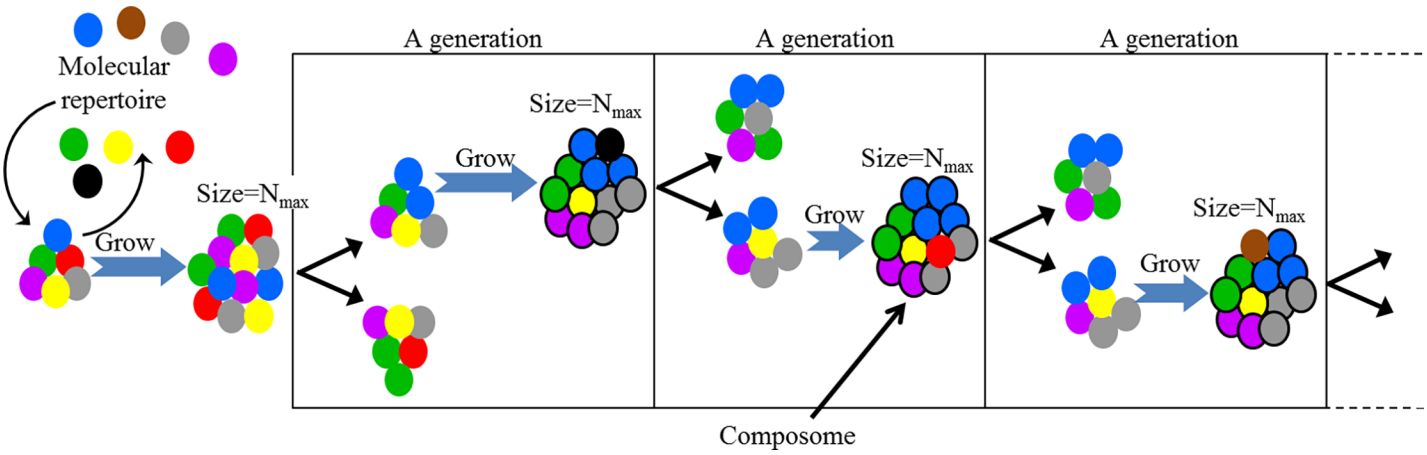

Fig 1. Schematics of GARD's dynamics. Different molecules types (represented by different colored circles) aggregate to form assemblies. Aggregation is biased by a matrix of chemical rates (β, Eq 1)). Once an assembly reaches a size-threshold (N_{max}) it splits, and the progeny then continues the growth-split cycles (generations). A composome is an assembly that has high average compositional similarity (see section: The GARD model) to its parent and to one of its children.

https://doi.org/10.1371/journal.pone.0192871.g001

As typical GARD simulations take a constant number of alphabet molecule types and a predefined assembly size, the total number of possible compositions is fixed [32] and the system is not permitted to show true open-ended evolution [43]. In 2010, perhaps the first rigor attempt at studying evolution in GARD was reported, in the sense of population responding to an external selection pressure [44]. Unfortunately, the study was based on a single instantiation of a random lognormal matrix, which hinders on the ability to draw conclusions from it. Moreover, the study employed parameters values very different than those typical used in GARD (i.e. small alphabet size and small assembly size), and the study did not designate compotype species as targets for selection. A later study considered a similar methodology for selection as the 2010 paper, and explored a large number of matrix instances and focused on compotypes as selection targets [**40**], asking whether compotypes change their frequency within a population as an outcome of external selection. The later study found that GARD systems can respond to selection (but not always), and that this selection response is more favourable when the matrix instance is highly mutualistic (i.e. when off-diagonal values are higher than diagonal values). A recent attempt to extend the 2010 paper by attempting to map GARD into the quasispecies formalism [45] presents an argument on GARD's putative limited evolvability. The paper failed however to designate compotypes as selection targets, even thought it was previously shown that only compotypes can be mapped into quasispecies [32], and used atypical GARD parameters.

Regardless of selection behaviour, the present paper asks whether the biological diversity that surrounds us would be different if the tape of life was to run again from the start [46–49] under similar initial conditions, and whether adaptations that lead to similar phenotypes follow a quantifiably repeatable route [50]. Some evidence for the convergent nature of evolution can be seen when two separated populations of *E. coli* evolved separately for many generations in identical environments achieved similar fitness [51], or when different populations of lizards from different nearby islands developed into similar ecomorphs independently [52]. Computer models have also been used to study this question [53–58].

In this paper we postulate that if evolutionary diversity is dominated by "invariants" rather than "decorations" then it should be possible to predict the outcome of the evolutionary process without actually waiting for it to happen. That is, it should be possible to predict which

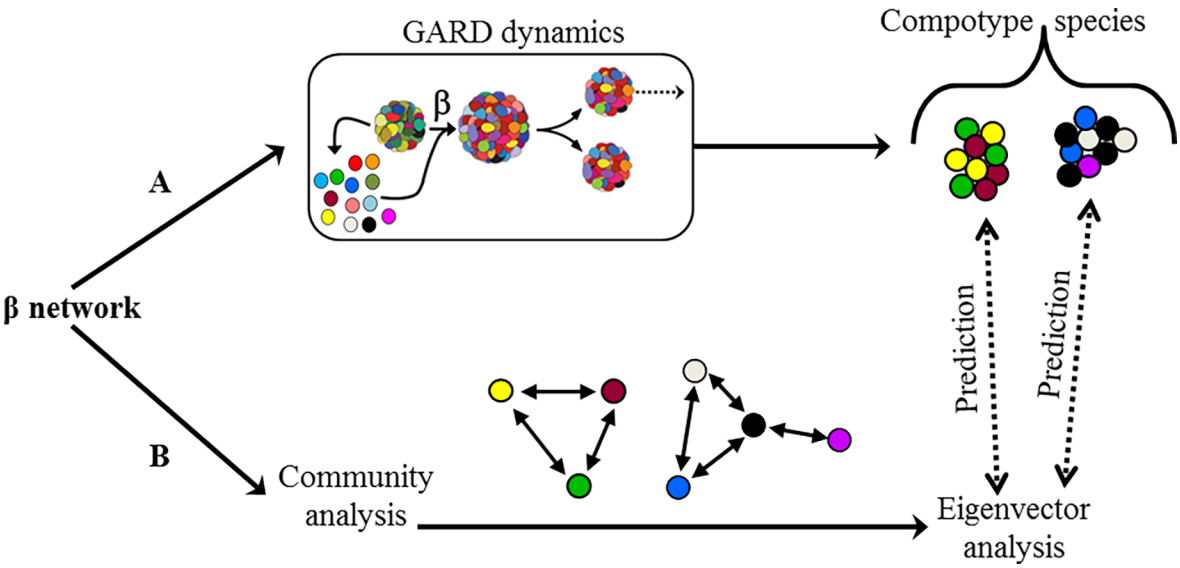

Fig 2. Overview of the algorithm developed in the present work. (**A**) A network (β) is employed in GARD simulations and the emerging compotype species are collected. (**B**) In parallel, the communities of β are analysed and collected. Finally, (A) and (B) are compared by using the ensemble of detected communities to predict compotypes.

https://doi.org/10.1371/journal.pone.0192871.g002

species will emerge. In the present paper, this translated to investigating the degree to which the emergence of GARD species, i.e. compotypes, can be analysed in terms of β's inner organization only (i.e. independent of the dynamics in GARD) (Fig 2). In order to do this, we analysed the community structure of β. Typically in a network representation, nodes symbolize entities (molecules, web pages, people, etc') and edges are relations between the entities (catalysis, hyperlinks, friendship, etc'). Communities are organizational features in many networks, and are generally defined as sets of nodes more densely interconnected between themselves than to other nodes in the network [59–61]. Communities detection algorithms allow revealing of essential internal network organization and typically detection algorithms try to optimise the ratio between the number of internal community to cross-communities edges across all communities simultaneously.

Network science is often fruitfully applied to decipher and understand complex systems, including food-webs [62], metabolic networks [63], genes networks [64], protein networks [65] and different social networks [66, 67]. Such applications of network science, together with previous linear algebra analysis of β [39] and of other networks [68, 69], motivate us to apply such analyses to our system, focusing on how the inner organization of a β affects the nature of observed compotypes species. Even though differences exist between replicating polymers and replicating catalytic networks [32], in both cases the model can be represented as a network [13] and encourages understanding how network's inner organisation affects the nature of observed species. We showed in [70] that one can predict the best simulation algorithms for systems and synthetic biology models by analysing their network structure. Further, different β's result in different GARD simulations giving rise to different compotype species provides additional motivation for our current study.

In this paper we use large scale simulations and data analysis of GARD simulations to demonstrate that communities' analysis allows us to "shortcut" expensive dynamical simulations of a (proto) evolutionary process and predict its invariants, namely, the set of species that can be expected to emerge from such a dynamical system.

Methods

The GARD model

GARD describes the growth and fission of a molecular assembly, typically assumed to consist of a large repertoire of amphiphilic molecules drawn from a repertoire of N_G molecular types [38, 40] (Fig 1). Molecules from the environment join an assembly and molecules within the assembly it can leave. Once the number of molecules in an assembly reaches a pre-defined size threshold (N_{max}), a random fission event takes place and produces two daughter assemblies of the same size ($N_{max}/2$) which can then repeat the growth-fission cycle (Fig 2 show a scheme of the model, adapted from [32]). This dynamic is described by a set of ordinary differential equations:

$$\frac{dn_i}{dt} = (k_f \rho_i N - k_b n_i)\left(1 + \sum_{j=1}^{N_G} \beta_{ij} \frac{n_j}{N}\right) \quad \text{Eq 1}$$

Where n_i is the current count of molecule type i in an assembly ($i = 1..N_G$), k_f and k_b are the basal forward and backward rate constants (assembly joining and leaving, respectively). ρ_i is the buffered environmental concentration and N is current assembly size ($N = \sum n_i$). β_{ij} is the rate-enhancement exerted by an assembly molecule of type j on incoming or outgoing molecule of type i.

β can be represented as an $N_G \times N_G$ adjacency matrix for a weighted-directed-asymmetric-network with N_G nodes and N_G^2 edges. Typically, β_{ij} values are drawn from a lognormal distribution [39, 71] (that is, the values $\ln(\beta_{ij})$ are normally distributed with mean = -4 and standard deviation = 4) where different β instances represent different potential environmental prebiotic chemistries [40]. Introducing negative β_{ij} values, i.e. inhibition, is expected to result in catalysis aswell via inhibition of inhibitor [40].

As mentioned previously, composomes are faithfully replicating assemblies, that is a composome is an assembly with high similarity to its predecessor and successor (typically compared when both assemblies are at size N_{max}). It is important to distinguish composome assemblies from non composomes (i.e. drifting assemblies), because the latter may appear spontaneously yet are incapable of transmitting compositional information (i.e. the specific ratio of different molecule types): that is, once a non composome assembly reaches the critical size triggering the fission event (N_{max}), its compositional information is not preserved in the daughter assemblies and hence is lost. Composomes are grouped into compotypes using k-means clustering algorithm based on compositional similarity as a distance measure (see section: Compotype-community assignment) by picking the k which give the highest silhouette [41]. A compotype is thus represented by a vector constituting the center of mass of all its member assemblies and is interpreted as a GARD species.

GARD simulations

The GARD model was run using a stochastic kinetic Monte Carlo simulation based on Gillespie's algorithm [72] using parameter values identical to those employed in previous studies [32, 40, 42]: $k_f = 10^{-2}$, $k_b = 10^{-4}$, $\rho_i = 10^{-2}$, $N_{max} = 10^2$ and $N_G = 10^2$, for 5,000 growth-split cycles (generations). Calculations were executed using MATLAB version R2015a. A large set of 10,000 GARD simulations was generated, all with the above parameters, and each with a different β, created by MATLAB's pseudorandom number generator with seeds 1–10,000. Each of these β's represents different chemistries that might lead to the emergence of one or more compotypes.

In the basic form employed for this paper, GARD was run in a single-lineage mode, where at each split event only one progeny (picked at random) is followed and the other one is

discarded (Fig 1). For each simulation under a given β, composomes where identified and clustered into compotypes.

Simulations give rise to the emergence of various compotype species as a result of the different chemistries represented by different β's. The number of compotypes observed in each simulation typically ranged from 1–6, with a total of 20,235 compotypes observed in 10,000 simulations performed (3 simulations out of those failed and were therefore discarded).

We provide the MATLAB code and datasets used in this work (see S1 File (Supporting Information) and reference [73]).

Community detection algorithms

A community detection algorithm was run on each β, and the list of nodes (molecule types) belonging to each community was recorded per each β. The three different algorithms used are: Louvain (MATLAB version) [74], Infomap (version 0.18.2) [75] and OSLOM (Order Statistics Local Optimization Method) (version 2.5) [76], with their default parameters.

Louvain is a heuristic method to find communities [74]. This method starts by assigning each node to its own community. Then, a node m is added to the community of node n only if this results in increased modularity value. m and n pairs are picked to give the highest increase. This is continued until no increase in modularity is gained by joining nodes. Next, a new network is created, whose nodes correspond to the previously found communities and whose edges are the respective sum of the previous edges between communities. This entire process is repeated until no further increase in modularity is possible.

Infomap is based on flow and encoding [75]. This method first simulates a random walk along the network, biased by the edges' weights. These random walks are then encoded into binary string in a way that would reflect how frequency adjacent nodes are visited, rather than create a maximally compressible binary string. This is done in a two-level description whereby a community of nodes where the walker has spent long periods of time receives unique code, but the nodes within a community receive non-unique codes that can be repeated in other communities' nodes. In other words, the random walk is efficiently encoded in a way in which important structures (communities) indeed retain unique codes.

OSLOM is finding clusters which are statistically significant with respect to a random network with similar characteristics as the actual network [76]. This method begins by randomly picking a node as the first community and additional nodes are added to this community if they are considered significance in the statistical sense. This is then repeated with other nodes until all communities are found.

Results and discussion

GARD tapes

In order to understand if convergent evolution is occurring under GARD dynamics, simulation-runs were repeated 10 times under a given β, with different random seeds (and hence initial assembly) each time. Each repeated run is regarded as a GARD "tape" (analogue to replaying the tape of life under the same chemistry (i.e. β)). The history of each tape was recorded (i.e. the content of each assembly) and compotypes were identified for each tape (i.e. k-means clustering). Fig 3 (panels A1-C1) show individual examples of GARD tapes (more examples are available at http://ico2s.org/data/extras/gard/). These panels show the content of assemblies from the different tapes, where different assemblies are plotted along the X axis and the N_G molecule types are of each assembly are given along the Y axis, with color representing the count of a molecule type in an assembly. While the detailed histories of various tapes under a given β are different, they generally show similar trends (invariants) represented by

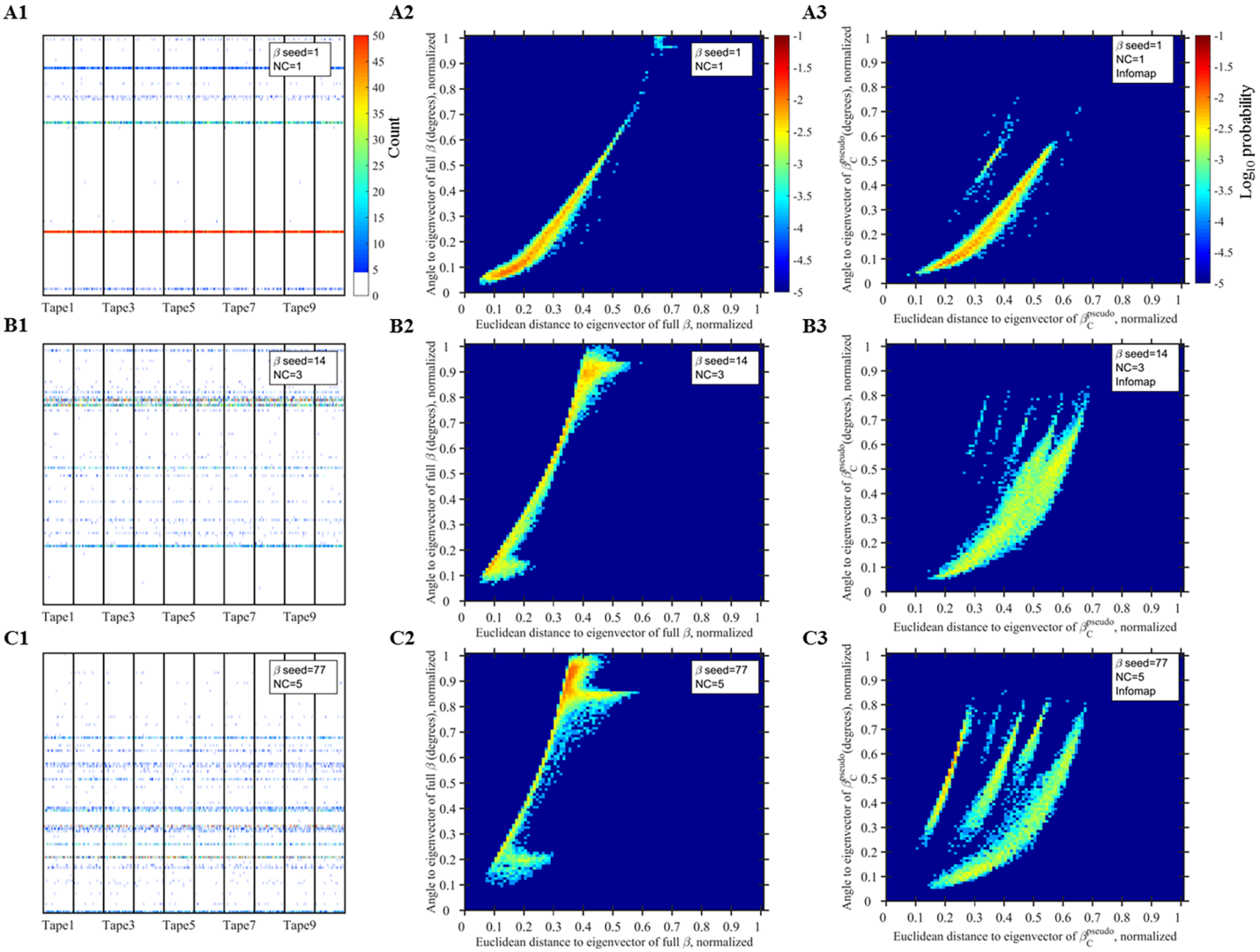

Fig 3. Examples of GARD simulations under different β's. (A1-C1) Histories of different tapes; For each tape, assemblies from different generations are plotted along the X axis, and color represents the counts of each of the N_G molecule types in each assembly (recorded at assembly size N_{max} (Fig 1). Tapes are separated by a vertical black line. For each tape, the first 1,000 assemblies are shown. Red color represents counts ≥ 50, and for brevity counts < 5 are colored white. (A2-C2) Density plots; For each assembly shown in panels (A1-C1), its Euclidean distance and angle vs. the eigenvector of the full-β was calculated (normalized for the maximum value between two assemblies, $N_{max}\sqrt{2}$ for distance and 90 degrees for angle). Color is normalized probability ($\log_{10}$ scale) of an assembly having a certain angle and distance. See section: Compotype-community assignment. (A3-C3) Same as (A2-C2), except for each assembly the distance and angle are calculated against the one eigenvector of β^* which has the lowest angle to this assembly. Number of Infomap communities detected is: 9 (A3), 7 (B3) and 6 (C3). Further examples are available at http://ico2s.org/data/extras/gard/ and [73].

https://doi.org/10.1371/journal.pone.0192871.g003

the horizontal lines. Further, different tapes from the same β exhibit the same number of compotypes (in 85% of cases studied for this part, see Fig A in S1 File), and, importantly, those compotypes are extremely similar between different tapes (Fig B in S1 File), signifying that GARD dynamics display convergent evolution. In other words, even if different GARD tapes portray different histories (decorations) under the same β, they give rise to very similar compotype species (invariants) and thus it becomes relevant to ascertain to what degree it is possible to predict the emergence of these species from the underlying chemistry alone, i.e., ignoring the dynamical process that generates the species. Because GARD exhibits convergent evolution, in the next sections only a single tape will be simulated per each β, but in return a large number of different β's will be employed.

Communities detection

This section presents how community detection algorithms were applied to the β network, and the next section presents how the detected communities were related to the emergence of species in the prebiotic evolution model (GARD) (Fig 2). In order to adequately compare a detected community to an observed compotype one needs to convert a community–which is a set of molecule types (nodes and their links)–to a composition, that is–the ratios between those molecule types. This composition can then be directly compared with the composition of a compotype. To detect communities within different β matrices, each of the three different algorithms used (Louvain [74], Infomap [75] and OSLOM [76]), were run on each β, and the list of nodes (molecule types) belonging to each community was recorded per each β.

Each of the three algorithms always detected several communities (>1) in each of the 10,000 different β's studied here (Fig 4 A and 4B). Louvain algorithm detected on average fewer communities than Infomap or OSLOM. Interestingly, both OSLOM and Infomap detected similar numbers of communities, even though OSLOM allows for overlaps (i.e. molecules belonging to more than 1 community). The latter suggests that a detection algorithm may sometime consider two overlapping communities as one, if overlaps are allowed. In GARD these overlaps are suggested to be the facilitators of species interconverting into each other–a phenomenon best seen in GARD populations [42].

Different simulations under different β's give rise to different compotypes, which calls for the search for a link between the inner structures of a β to the emerging compotypes in a simulation under this β. However, as the average number of communities detected in a β is higher than the average number of compotypes observed in a simulation under this β and no correlation between number of communities to number of compotype exists (Fig 4C), finding such link is not trivial. The next section will discuss a methodology for community–to–compotype assignment and prediction.

Compotype-community assignment

In order to perform such comparison, compotypes observed in each β-dependent simulation were collected (will sometime be referred to as original compotypes) and on the other hand the communities detected in each β were collected (respectively corresponding to (A) and (B) in Fig 2). Then, for each detected community in each β, a matrix β^* is created with elements β_{ij}^*:

$$\beta_{ij}^* = \begin{cases} \beta_{ij} & i \in C \text{ AND } j \in C \\ 0 & otherwise \end{cases} \quad \text{Eq 2}$$

Where C is the set of the indices of all nodes (molecule types) that belong to a community, i and j are nodes' indices and β_{ij} are elements of β (Eq 1). β^* has the same dimensions as β. That is– β^* is a sparser version of β matrix in which only pairs of molecule types that belong to a community can interact (all other rates are set to zero). This particular formulation of β^* was picked such that its eigenvectors will have the same dimensionality as the original compotypes. Next, linear algebra is used on β^*.

According to the Perron-Frobenius theorem a matrix such as β^* or β has a nondegenerate largest real eigenvalue with a corresponding eigenvector with all non-negative elements [77, 78]. Indeed, an eigenvector analysis on all the β^*'s and β's studied here showed that only a single non-negative eigenvector exists for each. It is tempting to consider an eigenvector with all non-negative elements as representing a molecular composition (as sometimes done [39, 77, 79]), homologue to a compotype. A vector with some negative elements, representing negative molecular counts or concentrations, by definition cannot represent molecular composition.

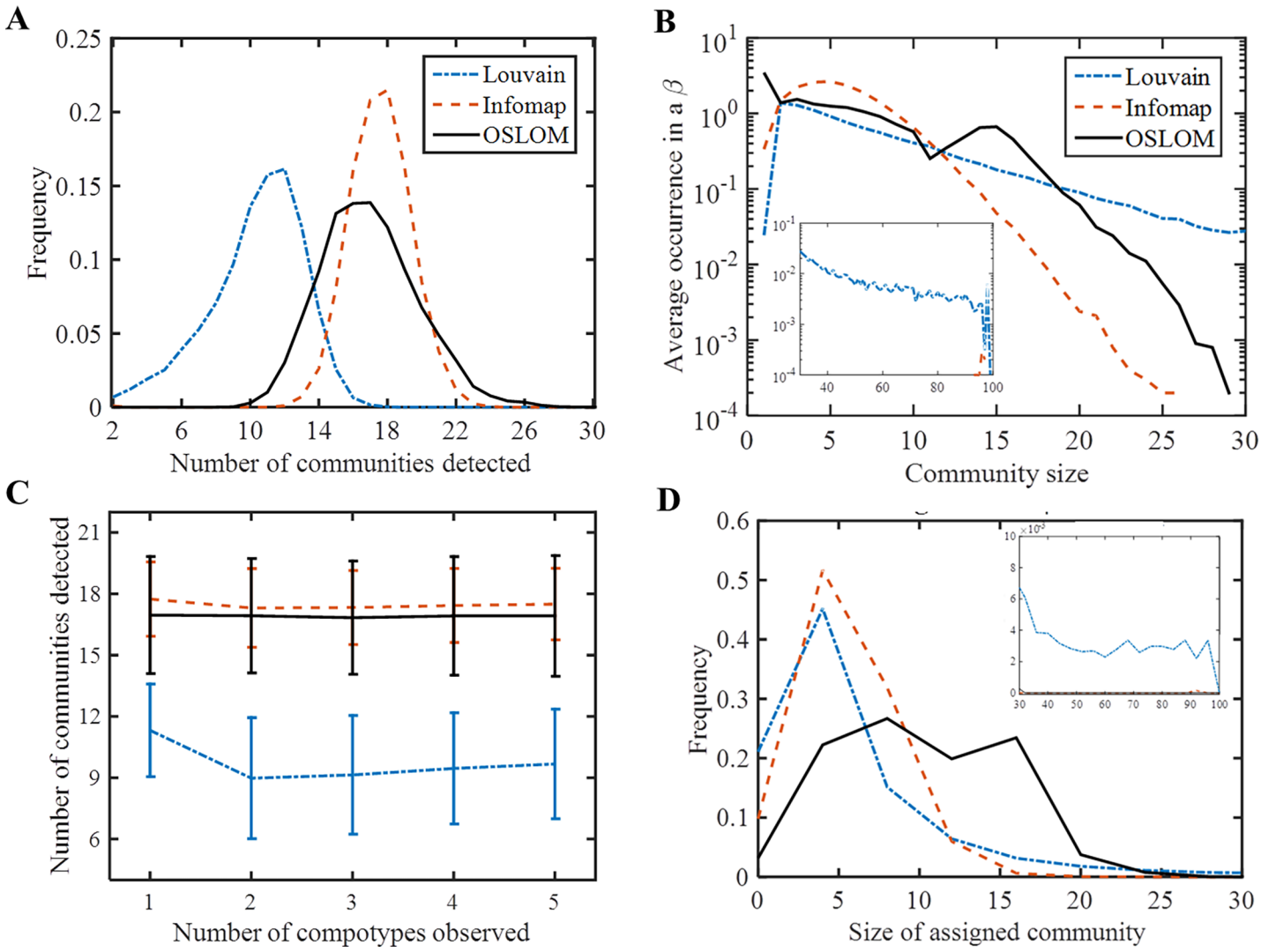

Fig 4. Communities in β's. (A) Histogram of total number of communities detected in each network, for the 3 algorithms. Frequency is given out of the 10^4 β's studied here. (B) Average occurrence of community-sizes. An occurrence of 10^{-4} means that this community-size appeared only in 1 β out of the 10,000 studied here and an occurrence of 1 means that on average each β has one community with this size. Insert show the occurrence of sizes > 30. (C) Average number of communities detected vs. number of observed compotype species shows no correlation. Vertical bars mark standard deviation. (D) Histogram of the size of assigned communities, when a community is assigned to a compotype based on eigenvector similarity (see section: Compotype-community assignment). Mean and standard deviation are given in Table 1.

https://doi.org/10.1371/journal.pone.0192871.g004

Because GARD simulations can exhibit more than 1 compotype (Fig 4C), it is unclear what is the relation between the single eigenvector of β to the observed compotypes and the same can be said about the communities. What follows presents a method to successfully predict the content (i.e. composition) of all compotypes observed in a simulation under a given β, given only the ensemble of communities of that β.

The Perron-Frobenius theorem was applied to all β^* and the eigenvectors were recorded. Exploring the role of communities in the actual GARD dynamics, the angle and distance between each assembly during a simulation to the eigenvector of the full-β and to the eigenvector of β^* were calculated (Fig 3 panels A2-C2 and A3-C3, respectively). Indeed, the assemblies show a lower angle to β^* than to β (see also Fig C in S1 File), symbolizing the significance of communities in analysing GARD's dynamics.

Each such eigenvector of β^* is compared with each compotype, using cosine of compotype vectors as typically applied in GARD studies [38, 41, 44, 80]:

$$H(V_1, V_2) = \cos(V_1, V_2) = \frac{V_1 \cdot V_2}{|V_1| \cdot |V_2|} \quad \text{Eq 3}$$

H measures how well an eigenvector matches a compotype's content (i.e. composition), where a value of 1.0 means identical compositions (i.e. one vector is the other vector multiplied by a positive scalar). Each compotype is then assigned with the community that give rise to the highest H.

Fig 5 shows, out of all the H values between the communities' eigenvectors and the original compotypes, the percentage of particularly high values (H>0.8). Full histograms are given in Fig D in S1 File. When multiple compotypes were observed in a simulation, the eigenvectors of β^* showed a high degree of similarity to all compotypes whereas the eigenvector of the full-β showed much lower similarity values (Table 1). Only in the limiting case, when only a single compotype is produced by the simulation, the eigenvector of the full-β showed high similarity to that compotype. Two-sample Kolmogorov-Smirnov tests were performed, with the null hypothesis that the similarities with respect to the full-β are from the same continuous distribution as the similarities with respect to β^*, against the alternative hypothesis that they are from different continuous distributions. The Kolmogorov-Smirnov tests were repeated for the cases of single and multiple compotypes, for each of the three community detection algorithms (that is– 6 tests in total). All the tests rejected the null hypothesis with alpha level that is

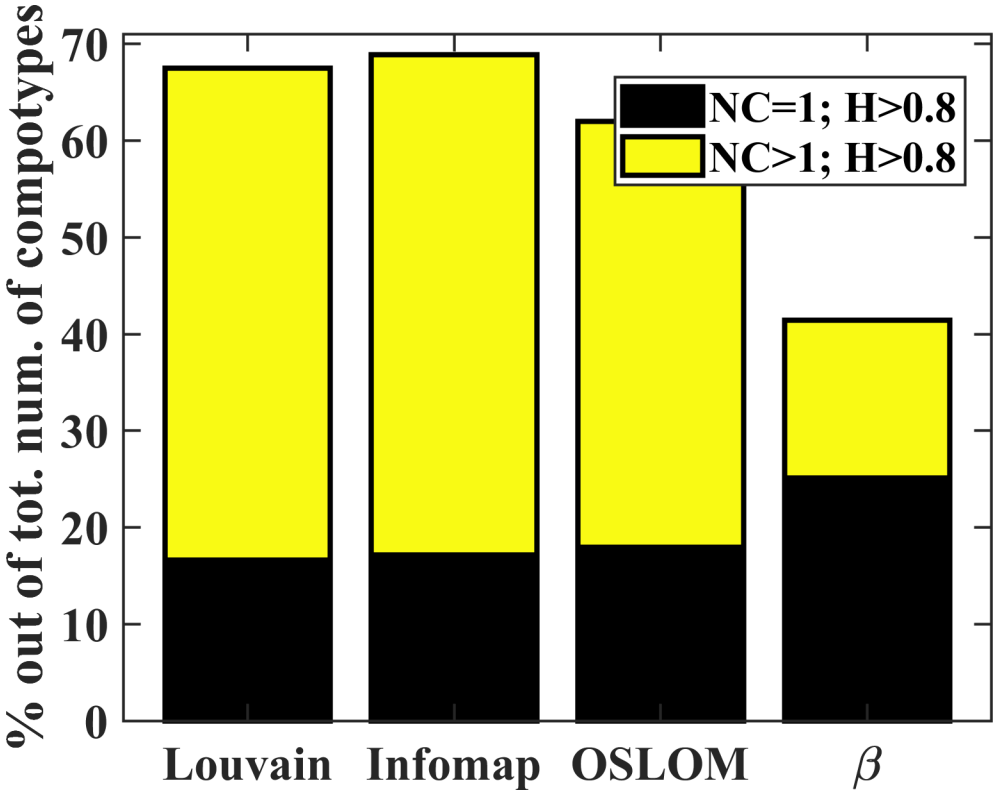

Fig 5. Bar plot of the percentage of high compositional similarity (H, Eq 3) when predicting compotypes using the eigenvectors of β^* (Eq 2) vs. full-β. Percentage is given out of the total number of compotypes observed under all β networks. Mean and standard deviation are given in Table 1 and full histograms are given in Fig D in S1 File.

https://doi.org/10.1371/journal.pone.0192871.g005

essentially zero. Further, when taking into account the overall dataset (that is, without distinguishing between cases with single or multiple compotypes), the majority of β^* showed substantial similarity to their original compotypes, with more than 60% of cases showing $H>0.8$ (Fig 5). The overall high degree of similarity achieved across all three community detection algorithms indicates that the communities are able to successfully predict the composition of compotypes, while the eigenvector of β may represent something else (see S1 File, section: On the eigenvector of the full-β). Thus, compotype species can be successfully predicted based only on the complex chemistry that is in a β. A test to ascertain whether a better community-to-compotype assignment and prediction could be achieved at random was performed. The test measured (for each community-to-compotype assignment) the probability of achieving higher H values by a random community–a community with the same size as the assigned community but with different molecule types. The test was repeated 10^3 times for each assigned community. The test showed that it is highly unlikely to achieve better H values by random community assignment (Table 1, and Fig E in S1 File).

Finally, it is important to verify whether indeed β^* represents a meaningful chemistry that can give rise to a compotype species under GARD's stochastic dynamics (Eq 1). To this end, GARD simulations were repeated with exactly the same parameters (see section: GARD simulations), and with β^* for each assigned community rather than with the full β. Compotype identification in the new simulations was performed exactly as before (i.e., k-means clustering) and the compositional similarity to the original compotype was calculated (Fig 6 'Original'). A high similarity to the original compotype was always obtained, corroborating the community detection algorithms ability to detect the communities which serve as the 'invariant content' of GARD's compotypes. In [81], the authors analysed stochastic Kauffman-like dynamics via the introduction of a temporal-window in order to determine which part of their reaction network is currently active, however, the novelty of the present paper is in enabling to make such determination *a-priori* based on the network topology.

As presently it is impossible to determine *a priori* the number of compotype species that will be observed, the algorithm for compotype-community assignment presented above is required to address all compotypes (however, it was previously shown that having an excess of mutual-interactions over self-interactions in β (i.e. β_{ij} over β_{ii}) is a necessary but insufficient condition for a high number of compotypes [40]).

On the nature of non-assigned communities

Lastly, it is asked why some communities successfully predict compotypes while other communities do not, and are there differences between those communities. It was previously suggested that compotype dynamics are somehow related to the compartments formed by high β_{ij} values [44]. The morphology of the communities assigned to compotypes seems to be different

Table 1. Statistics related to communities and compotypes.

		Louvain	Infomap	OSLOM	β
NC = 1	Mean	0.829±0.157	0.839±0.149	0.845±0.158	0.975±0.540
NC>1	Mean	0.797±0.217	0.823±0.171	0.747±0.254	0.624±0.269
Overall	Mean	0.805±0.204	0.827±0.166	0.773±0.237	0.716±0.280
Probability of a better similarity at random		0.0257±.135	0.00388±0.0156	0.0137±0.0399	
Size of assigned community		9±14	6±3	10±5	

Mean, standard deviation and percentage of dataset achieving high similarity between the eigenvectors of β^* and full-β and the original compotypes (NC = 1, cases when single compotype observed; NC>1, cases when multiple compotypes observed; Overall, the entire dataset), for the three algorithms (Fig D in S1 File).

https://doi.org/10.1371/journal.pone.0192871.t001

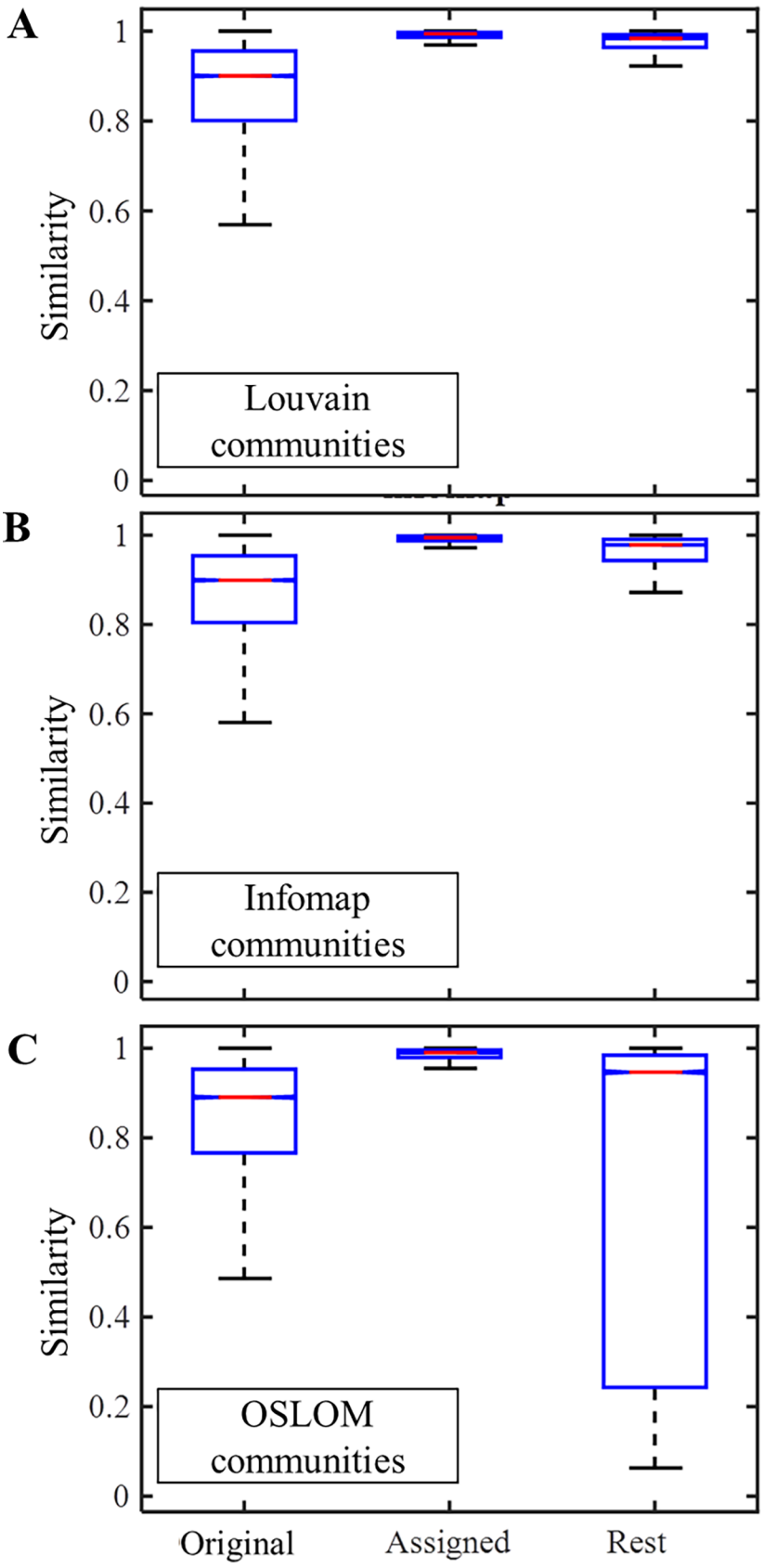
A
B
C
Similarity
1
0.8
0.6
0.4
0.2
0
Louvain communities
Infomap communities
OSLOM communities
Original
Assigned
Rest

Fig 6. Box plots of compositional similarity, for the three community detection algorithms (Louvain, top; Infomap, middle; OSLOM, bottom). Similarity was measured in three cases: '*Original*', when comparing the original compotype observed vs. the one in GARD under β^* of its assigned community; '*Assigned*', when comparing the compotype observed in a GARD simulation with β^* of its assigned community to the eigenvector of β^*; '*Rest*', analogue to 'Assigned', only with communities that were not assigned to original compotypes β^*. Mean and standard deviations for '*Original*' respectively are: 0.849±0.165, 0.856±0.145 and 0.813±0.216.

https://doi.org/10.1371/journal.pone.0192871.g006

than that of those which were not assigned (Fig 7), which may begin to point to the nature of differences between the assigned and non-assigned detected communities. Additionally, the similarity between the eigenvector of β^* and the compotype from GARD under β^* was calculated, both for the assigned and non-assigned communities. It was found that this similarity is much higher for the assigned communities than for the non-assigned (Fig 6 'Assigned' and 'Rest'). This last result suggests that the dynamics of the non-assigned communities is fundamentally different than that of the assigned ones, in the sense that the former are less likely to exhibit faithful replication. An ongoing investigation is on its way to further understand those differences, which may prove critical for reverse-engineering, i.e. the design of a β network that give rise to specific and desired compotypes dynamics.

Conclusions

The GARD model performs biased and far from equilibrium random walks on a network that has previously been linked to pre-biotic evolutionary dynamics. Via community analyses, we were able to bypass the dynamic trajectories of the stochastic simulator and use the ensemble of detected communities to predict the emergence of (proto) species of this system as well as their invariant content. Interestingly, the morphology of assigned communities is different than that of non-assigned ones, which deserve further scrutiny in order to understand the nature of this difference, how the various topological characteristics affect dynamics as well as the precise role of those un-assigned communities.

We have used the eigenvector of β^* to predict compotypes and corroborated by performing GARD dynamics under β^*, to find that GARD-dynamics approach gives rise to a compotype more similar to the original one (the original compotype observed under the full-β). In other words: using β^*, GARD-dynamics are 'closer to the truth'. This is both non-intuitive and interesting, because the eigenvector approach does not employ GARD's stochastic dynamics, where the latter are expected to introduce some variation in the compotype content. If we treat the observation of species in GARD's dynamics as the ground truth–analogous to how species are observed in nature–then this points that the theoretical prediction using the eigenvector is imperfect (but still very good!), probably because the eigenvector method takes into account only β and not the full physio-chemical details of the GARD model, such as the reversibility of assembly-joining.

For tractability, the present manuscript kept to the definition and identification of compotype species as they have traditionally been used in GARD and lipid world literature [32, 40–42, 80]. We would like to argue in favour of rethinking species identification, as follows. We speculate that the un-assigned communities represent either assemblies that are unable to faithfully replicate or compotype species that are very rare. The latter may require an even larger scale simulation analysis than the one we have done here involving more runs and longer simulation times before these rare species could be observed. Any species identification algorithm developed must, critically, acknowledge faithful replication. As presently it is impossible to determine a-priori the number of compotype species that will be observed in a simulation under a given β network, we are in the process of extending this current paper in order to precisely predict the expected number of compotype species under a given β without running

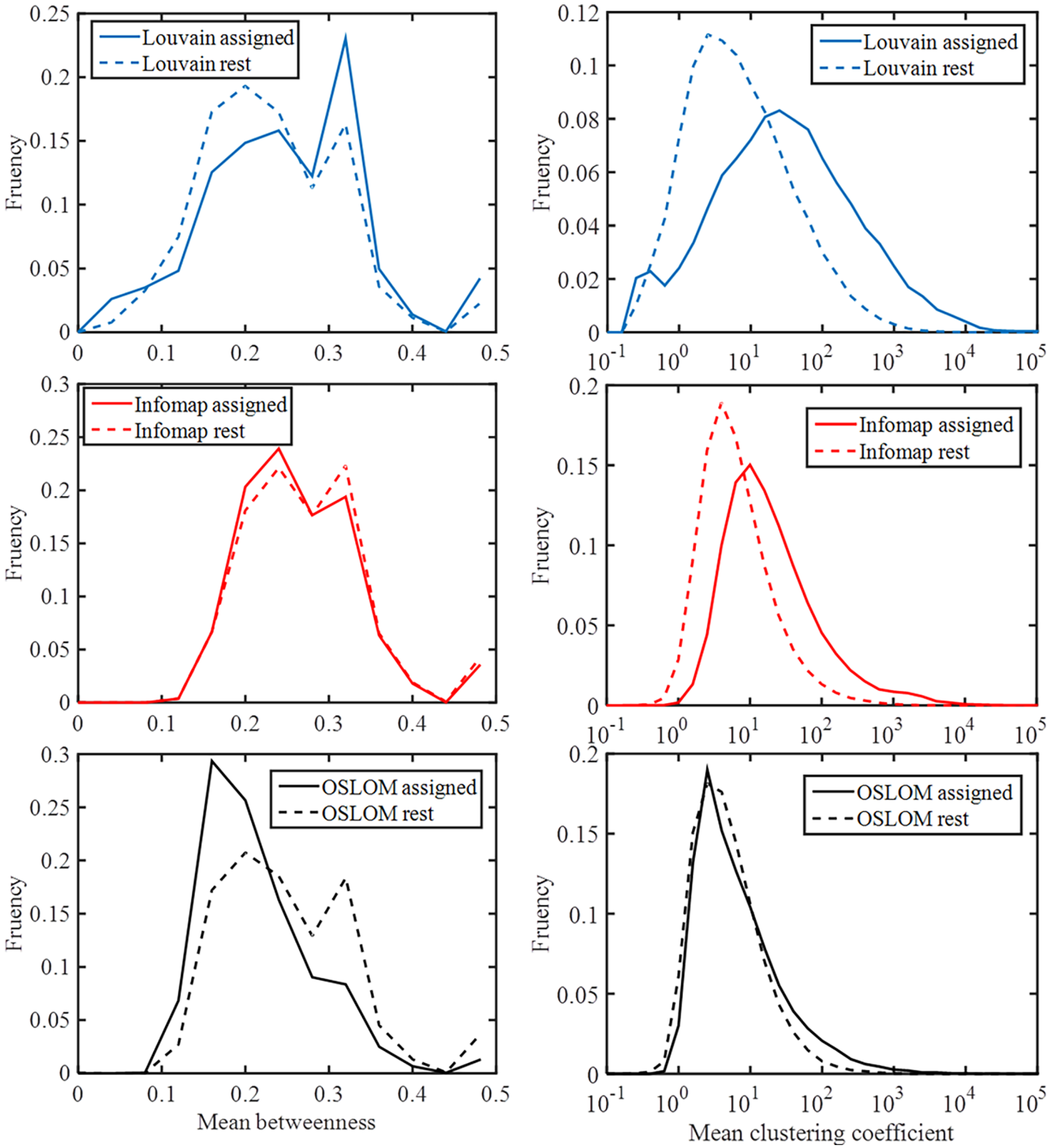

Fig 7. Network topology for assigned communities and rest, for the three community-detection algorithms (Louvain, top; Infomap, middle, OSLOM, bottom). (left) Node-betweenness-centrality [82], normalized by dividing with (n-1)*(n-2), where n is number of nodes in a community. (right) Clustering-coefficient [83]. Parameters were calculated using [84].

https://doi.org/10.1371/journal.pone.0192871.g007

simulations. The community count provides an upper limit for the species count, and the community eigenvectors, even if somewhat numerous, still strongly narrows the search for compotypes.

Our heuristic approach gave very similar results among all three community-detection-algorithms we used, thus providing robustness to our findings. Future extension of this work will apply the species-prediction-algorithm developed herein on multiple dynamical models and their emergent species (or species equivalent), as well as address larger networks which is more realistic, in order to address the generality of the algorithm presented here.

Supporting information

S1 File. Supplementary data and figures associated with this article.
(PDF) https://doi.org/10.1371/journal.pone.0192871.s001

Acknowledgments

We thank Harold Fellermann and Pawel Widera for discussions, and Jean-Loup Guillaume for making available Louvain's MATLAB code.

Author Contributions

Conceptualization: Omer Markovitch, Natalio Krasnogor.

Data curation: Omer Markovitch.

Formal analysis: Omer Markovitch.

Funding acquisition: Natalio Krasnogor.

Investigation: Omer Markovitch.

Methodology: Omer Markovitch, Natalio Krasnogor.

Project administration: Natalio Krasnogor.

Resources: Natalio Krasnogor.

Software: Omer Markovitch.

Supervision: Natalio Krasnogor.

Validation: Omer Markovitch.

Visualization: Omer Markovitch, Natalio Krasnogor.

Writing – original draft: Omer Markovitch, Natalio Krasnogor.

Writing – review & editing: Omer Markovitch, Natalio Krasnogor.

References

1. Rasmussen S. Protocells: Mit Press; 2009.
2. Dyson F. Origins of life: Cambridge University Press; 1999.
3. Powner MW, Sutherland JD. Prebiotic chemistry: a new modus operandi. Philos T R Soc B. 2011; 366 (1580):2870–7. https://doi.org/10.1098/rstb.2011.0134 WOS:000294993100003. PMID: 21930577
4. Walker SI, Davies PCW. The algorithmic origins of life. J R Soc Interface. 2013; 10(79). https://doi.org/10.1098/Rsif.2012.0869 WOS:000331118500006. PMID: 23235265
5. Cronin L, Krasnogor N, Davis BG, Alexander C, Robertson N, Steinke JH, et al. The imitation game—a computational chemical approach to recognizing life. Nature biotechnology. 2006; 24(10):1203–6. https://doi.org/10.1038/nbt1006-1203 PMID: 17033651

6. Benner SA. Defining life. Astrobiology. 2010; 10(10):1021–30. https://doi.org/10.1089/ast.2010.0524 PMID: 21162682
7. Trifonov EN. Vocabulary of Definitions of Life Suggests a Definition. J Biomol Struct Dyn. 2011; 29 (2):259–66. https://doi.org/10.1080/073911011010524992 ISI:000294884000004. PMID: 21875147
8. Oparin AI. Origin and evolution of metabolism. Comparative biochemistry and physiology. 1962; 4:371–7. Epub 1962/10/01. https://doi.org/10.1007/BF01218509 PMID: 13940205
9. Anet FA. The place of metabolism in the origin of life. Current Opinion in Chemical Biology. 2004; 8 (6):654–9. https://doi.org/10.1016/j.cbpa.2004.10.005ISI:000225782300014. PMID: 15556411
10. Szathmary E, Santos M, Fernando C. Evolutionary potential and requirements for minimal protocells. Top Curr Chem. 2005; 259:167–211. https://doi.org/10.1007/tcc001 WOS:000234567400005.
11. Shapiro R. Small molecule interactions were central to the origin of life. Quarterly Review of Biology. 2006; 81(2):105–25. https://doi.org/10.1086/506024 ISI:000237887600001. PMID: 16776061
12. Cleaves H. Prebiotic Chemistry: Geochemical Context and Reaction Screening. Life. 2013; 3(2):331. https://doi.org/10.3390/life3020331 PMID: 25369745
13. Kauffman SA. The origins of order: Self organization and selection in evolution: Oxford university press; 1993.
14. Szathmary E, Smith JM. From replicators to reproducers: the first major transitions leading to life. J Theor Biol. 1997; 187(4):555–71. https://doi.org/10.1006/jtbi.1996.0389 WOS:A1997XR80400010. PMID: 9299299
15. Chen IA, Walde P. From Self-Assembled Vesicles to Protocells. Csh Perspect Biol. 2010; 2(7). https://doi.org/10.1101/cshperspect.a002170 WOS:000279883100009. PMID: 20519344
16. Bernhardt HS. The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others). Biology direct. 2012; 7:23. Epub 2012/07/17. https://doi.org/10.1186/1745-6150-7-23 PubMed Central PMCID: PMC3495036. PMID: 22793875
17. Gesteland FR, Cech RT, Atkins FJ. The RNA world. Cold Spring: Cold Spring Harbor Laboratory; 1999. 709 p.
18. Gilbert W. Origin of Life—the RNA World. Nature. 1986; 319(6055):618–. https://doi.org/10.1038/319618a0 ISI:A1986A079600021.
19. Joyce GF. The antiquity of RNA-based evolution. Nature. 2002; 418(6894):214–21. https://doi.org/10.1038/418214a ISI:000176710400049. PMID: 12110897
20. Orgel LE. Prebiotic chemistry and the origin of the RNA world. Critical Reviews in Biochemistry and Molecular Biology. 2004; 39(2):99–123. https://doi.org/10.1080/10409230490460765 ISI:000222588800002. PMID: 15217990
21. Luisi PL, Walde P, Oberholzer T. Lipid vesicles as possible intermediates in the origin of life. Current Opinion in Colloid & Interface Science. 1999; 4(1):33–9. https://doi.org/10.1016/S1359-0294(99)00012-6 ISI:000080941800005.
22. Segre D, Ben-Eli D, Deamer DW, Lancet D. The lipid world. Origins Life Evol B. 2001; 31(1–2):119–45. https://doi.org/10.1023/A:1006746807104 ISI:000167737900008.
23. Aono M, Kitadai N, Oono Y. A Principled Approach to the Origin Problem. Origins of life and evolution of the biosphere: the journal of the International Society for the Study of the Origin of Life. 2015; 45 (3):327–38. https://doi.org/10.1007/s11084-015-9444-3 PMID: 26177711; PubMed Central PMCID: PMC4510921.
24. Hanczyc MM, Fujikawa SM, Szostak JW. Experimental models of primitive cellular compartments: encapsulation, growth, and division. Science. 2003; 302(5645):618–22. Epub 2003/10/25. https://doi.org/10.1126/science.1089904 PMID: 14576428.
25. Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell. 1982; 31 (1):147–57. Epub 1982/11/01. https://doi.org/10.1016/0092-8674(82)90414-7 PMID: 6297745.
26. Lincoln TA, Joyce GF. Self-Sustained Replication of an RNA Enzyme. Science. 2009; 323(5918):1229–32. https://doi.org/10.1126/science.1167856 ISI:000263687600043. PMID: 19131595
27. Hayden EJ, Lehman N. Self-assembly of a group I intron from inactive oligonucleotide fragments. Chem Biol. 2006; 13(8):909–18. https://doi.org/10.1016/j.chembiol.2006.06.014 ISI:000240329800015. PMID: 16931340
28. Vaidya N, Manapat ML, Chen IA, Xulvi-Brunet R, Hayden EJ, Lehman N. Spontaneous network formation among cooperative RNA replicators. Nature. 2012; 491(7422):72–7. https://doi.org/10.1038/nature11549 ISI:000310434500032. PMID: 23075853
29. Lazcano A. Historical Development of Origins Research. Cold Spring Harb Perspect Biol. 2010; 2(11). https://doi.org/10.1101/cshperspect.a002089 ISI:000283646900001. PMID: 20534710

30. Oparin AI. Evolution of Concepts of Origin of Life, 1924–1974. Origins Life Evol B. 1976; 7(1):3–8. https://doi.org/10.1007/Bf01218509 WOS:A1976CE36200001.

31. Segre D, Lancet D. Composing life. Embo Rep. 2000; 1(3):217–22. https://doi.org/10.1093/embo-reports/kvd063 ISI:000165766200009. PMID: 11256602

32. Gross R, Fouxon I, Lancet D, Markovitch O. Quasispecies in population of compositional assemblies. BMC evolutionary biology. 2014; 14:265. https://doi.org/10.1186/s12862-014-0265-1 PMID: 25547629; PubMed Central PMCID: PMC4357159.

33. Maurer SE, Deamer DW, Boncella JM, Monnard PA. Chemical Evolution of Amphiphiles: Glycerol Monoacyl Derivatives Stabilize Plausible Prebiotic Membranes. Astrobiology. 2009; 9(10):979–87. https://doi.org/10.1089/ast.2009.0384 WOS:000273181200006. PMID: 20041750

34. Vequi-Suplicy CC, Riske KA, Knorr RL, Dimova R. Vesicles with charged domains. Bba-Biomembranes. 2010; 1798(7):1338–47. https://doi.org/10.1016/j.bbamem.2009.12.023 WOS:000279365000008. PMID: 20044978

35. Theis M, Gazzola G, Forlin M, Poli I, Hanczyc MM, Bedau M. Optimal formulation of complex chemical systems with a genetic algorithm. ECCS06 online Proceedings (P193). Oxford; 2009.

36. Gutierrez JMP, Hinkley T, Taylor JW, Yanev K, Cronin L. Evolution of oil droplets in a chemorobotic platform. Nat Commun. 2014;5. https://doi.org/10.1038/Ncomms6571 WOS:000347223500001. PMID: 25482304

37. Shirt-Ediss B, Sole RV, Ruiz-Mirazo K. Emergent chemical behavior in variable-volume protocells. Life (Basel). 2015; 5(1):181–211. https://doi.org/10.3390/life5010181 PMID: 25590570; PubMed Central PMCID: PMC4390847.

38. Segre D, Ben-Eli D, Lancet D. Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. Proceedings of the National Academy of Sciences of the United States of America. 2000; 97(8):4112–7. https://doi.org/10.1073/pnas.97.8.4112 PMID: 10760281; PubMed Central PMCID: PMC18166.

39. Segre D, Shenhav B, Kafri R, Lancet D. The molecular roots of compositional inheritance. J Theor Biol. 2001; 213(3):481–91. https://doi.org/10.1006/jtbi.2001.2440 PMID: 11735293.

40. Markovitch O, Lancet D. Excess mutual catalysis is required for effective evolvability. Artificial life. 2012; 18(3):243–66. https://doi.org/10.1162/artl_a_00064 PMID: 22662913.

41. Shenhav B, Oz A, Lancet D. Coevolution of compositional protocells and their environment. Philos T R Soc B. 2007; 362(1486):1813–9. https://doi.org/10.1098/rstb.2007.2073 ISI:000249516700009. PMID: 17510019

42. Markovitch O, Lancet D. Multispecies population dynamics of prebiotic compositional assemblies. J Theor Biol. 2014; 357:26–34. https://doi.org/10.1016/j.jtbi.2014.05.005 PMID: 24831416.

43. Markovitch O, Sorek D, Lui LT, Lancet D, Krasnogor N. Is there an optimal level of open-endedness in prebiotic evolution? Origins of life and evolution of the biosphere: the journal of the International Society for the Study of the Origin of Life. 2012; 42(5):469–73; discussion 74. https://doi.org/10.1007/s11084-012-9309-y PMID: 23114973.

44. Vasas V, Szathmáry E, Santos M. Lack of evolvability in self-sustaining autocatalytic networks constraints metabolism-first scenarios for the origin of life. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107(4):1470–5. https://doi.org/10.1073/pnas.0912628107 ISI:000273974600045. PMID: 20080693

45. Vasas V, Fernando C, Szilagyi A, Zachar I, Santos M, Szathmary E. Primordial evolvability: Impasses and challenges. J Theor Biol. 2015; 381:29–38. https://doi.org/10.1016/j.jtbi.2015.06.047 PMID: 26165453.

46. Gould. Wonderful Life: The Burgess Shale and the Nature of History (Book). Library Journal. 1989; 114 (14):214–.

47. Kauffman S. At home in the universe: The search for the laws of self-organization and complexity: Oxford university press; 1996.

48. Beatty J. Replaying life's tape. The Journal of philosophy. 2006; 103(7):336–62. https://doi.org/10.5840/jphil2006103716

49. Orgogozo V. Replaying the tape of life in the twenty-first century. Interface focus. 2015; 5(6):20150057. https://doi.org/10.1098/rsfs.2015.0057 PMID: 26640652; PubMed Central PMCID: PMC4633862.

50. Lobkovsky AE, Koonin EV. Replaying the tape of life: quantification of the predictability of evolution. Frontiers in genetics. 2012; 3:246. https://doi.org/10.3389/fgene.2012.00246 PMID: 23226153; PubMed Central PMCID: PMC3509945.

51. Travisano M, Mongold JA, Bennett AF, Lenski RE. Experimental tests of the roles of adaptation, chance, and history in evolution. Science. 1995; 267(5194):87–90. PMID: 7809610.

52. Losos JB, Jackman TR, Larson A, Queiroz K, Rodriguez-Schettino L. Contingency and determinism in replicated adaptive radiations of island lizards. Science. 1998; 279(5359):2115–8. PMID: 9516114.

53. Fontana W, Buss LW. What would be conserved if "the tape were played twice"? Proceedings of the National Academy of Sciences of the United States of America. 1994; 91(2):757–61. PMID: 8290596; PubMed Central PMCID: PMC43028.

54. Szathmary E. A classification of replicators and lambda-calculus models of biological organization. Proceedings Biological sciences / The Royal Society. 1995; 260(1359):279–86. https://doi.org/10.1098/rspb.1995.0092 PMID: 7630896.

55. Taylor T, Hallam J. Replaying the tape: An investigation into the role of contingency in evolution. From Anim Animat. 1998:256–65. WOS:000075924900029.

56. Bedau M. The scientific and philosophical scope of artificial life. Leonardo. 2002; 35(4):395–400. https://doi.org/10.1162/002409402760181196 WOS:000177435800010.

57. Wagenaar DA, Adami C. Influence of chance, history, and adaptation on digital evolution. Artificial life. 2004; 10(2):181–90. https://doi.org/10.1162/106454604773563603 WOS:000220379900007. PMID: 15107230

58. Missa O, Dytham C, Morlon H. Understanding how biodiversity unfolds through time under neutral theory. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2016; 371 (1691). https://doi.org/10.1098/rstb.2015.0226 PMID: 26977066; PubMed Central PMCID: PMC4810819.

59. Fortunato S. Community detection in graphs. Physics Reports. 2010; 486(3):75–174. https://doi.org/10.1016/j.physrep.2009.11.002

60. Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms. Physical review E. 2008; 78(4):046110.

61. Kim Y, Son SW, Jeong H. Finding communities in directed networks. Physical review E, Statistical, nonlinear, and soft matter physics. 2010; 81(1 Pt 2):016103. https://doi.org/10.1103/PhysRevE.81.016103 PMID: 20365428.

62. Sole RV, Montoya JM. Complexity and fragility in ecological networks. Proceedings Biological sciences / The Royal Society. 2001; 268(1480):2039–45. https://doi.org/10.1098/rspb.2001.1767 PMID: 11571051; PubMed Central PMCID: PMC1088846.

63. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. The large-scale organization of metabolic networks. Nature. 2000; 407(6804):651–4. https://doi.org/10.1038/35036627 PMID: 11034217.

64. Bassel GW, Lan H, Glaab E, Gibbs DJ, Gerjets T, Krasnogor N, et al. Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108(23):9709–14. https://doi.org/10.1073/pnas.1100958108 PMID: 21593420; PubMed Central PMCID: PMC3111290.

65. Glaab E, Baudot A, Krasnogor N, Valencia A. Extending pathways and processes using molecular interaction networks to analyse cancer genome data. BMC bioinformatics. 2010; 11:597. https://doi.org/10.1186/1471-2105-11-597 PMID: 21144022; PubMed Central PMCID: PMC3017081.

66. Ferrer ICR, Sole RV. The small world of human language. Proceedings Biological sciences / The Royal Society. 2001; 268(1482):2261–5. https://doi.org/10.1098/rspb.2001.1800 PMID: 11674874; PubMed Central PMCID: PMC1088874.

67. Barabasi AL, Jeong H, Neda Z, Ravasz E, Schubert A, Vicsek T. Evolution of the social network of scientific collaborations. Physica A. 2002; 311(3–4):590–614. https://doi.org/10.1016/S0378-4371(02)00736-7 WOS:000177271100023.

68. Schilling CH, Palsson BO. The underlying pathway structure of biochemical reaction networks. Proceedings of the National Academy of Sciences. 1998; 95(8):4193–8.

69. Jain S, Krishna S. Autocatalytic Sets and the Growth of Complexity in an Evolutionary Model. Physical Review Letters. 1998; 81(25):5684–7.

70. Sanassy D, Widera P, Krasnogor N. Meta-stochastic simulation of biochemical models for systems and synthetic biology. ACS synthetic biology. 2015; 4(1):39–47. https://doi.org/10.1021/sb5001406 PMID: 25152014.

71. Limpert E, Stahel WA, Abbt M. Log-normal Distributions across the Sciences: Keys and Clues On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question. BioScience. 2001; 51(5):341–52. https://doi.org/10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2

72. Gillespie DT. General Method for Numerically Simulating Stochastic Time Evolution of Coupled Chemical-Reactions. J Comput Phys. 1976; 22(4):403–34. https://doi.org/10.1016/0021-9991(76)90041-3 WOS:A1976CQ87900001.

73. Markovitch O, Krasnogor N. Accompanying dataset for: Predicting Species Emergence in Simulated Complex Pre-Biotic Networks. Zenodo. 2016. https://doi.org/10.5281/zenodo.56534

74. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. 2008; 2008(10):P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

75. Rosvall M, Axelsson D, Bergstrom CT. The map equation. The European Physical Journal Special Topics. 2010; 178(1):13–23. https://doi.org/10.1140/epjst/e2010-01179-1

76. Lancichinetti A, Radicchi F, Ramasco JJ, Fortunato S. Finding statistically significant communities in networks. PloS one. 2011; 6(4):e18961. https://doi.org/10.1371/journal.pone.0018961 PMID: 21559480; PubMed Central PMCID: PMC3084717.

77. Eigen M, Mccaskill J, Schuster P. Molecular Quasi-Species. J Phys Chem-Us. 1988; 92(24):6881–91. https://doi.org/10.1021/J100335a010 ISI:A1988R227300010.

78. Meyer CD. Matrix analysis and applied linear algebra. Philadelphia: Society for Industrial and Applied Mathematics; 2000. 718 p.

79. Virgo N, Ikegami T, McGregor S. Complex Autocatalysis in Simple Chemistries. Artificial life. 2016:1–15.

80. Armstrong DL, Markovitch O, Zidovetzki R, Lancet D. Replication of simulated prebiotic amphiphile vesicles controlled by experimental lipid physicochemical properties. Phys Biol. 2011; 8(6). https://doi.org/10.1088/1478-3975/8/6/066001 WOS:000298181900003. PMID: 21946049

81. Filisetti A, Graudenzi A, Serra R, Villani M, Füchslin RM, Packard N, et al. A stochastic model of autocatalytic reaction networks. Theory in Biosciences. 2012; 131(2):85–93. https://doi.org/10.1007/s12064-011-0136-x PMID: 21979857

82. Brandes U. A faster algorithm for betweenness centrality*. Journal of Mathematical Sociology. 2001; 25(2):163–77. https://doi.org/10.1080/0022250X.2001.9990249

83. Fagiolo G. Clustering in complex directed networks. Physical review E, Statistical, nonlinear, and soft matter physics. 2007; 76(2 Pt 2):026107. https://doi.org/10.1103/PhysRevE.76.026107 PMID: 17930104.

84. Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. NeuroImage. 2010; 52(3):1059–69. https://doi.org/10.1016/j.neuroimage.2009.10.003 PMID: 19819337.